Robust Energy Minimization for BRDF-Invariant Shape from Light Fields: Supplementary Material

Zhengqin Li Zexiang Xu Ravi Ramamoorthi Manmohan Chandraker
University of California, San Diego
{zhl378, zex014, ravir, mkchandraker}@eng.ucsd.edu

The supplementary material is divided into three parts. In Section 1, we present a detailed comparison on the effectiveness of various terms of our energy function using synthetic data. In Section 2, we show more mesh reconstructions with real data using a single light-field image. In Section 3, we present detailed derivations of our optimization framework for depth reconstruction and the BRDF-invariant term.

1. Ablation Study with Synthetic Data

Figure 1 shows qualitative comparisons of surface depths estimated from a synthetic light-field image, with different variations of our energy function. We also show the relative performance of two related works – one that uses a Lambertian shading model instead of BRDF-invariance [2] and another that uses BRDF-invariance, but without a robust variational minimization framework and without achieving a balance with traditional gradient-based multiview stereo [6]. Quantitative comparisons based on mean square error are demonstrated in Table 1.

The first column shows the input image while the second column is the ground truth depth map. The depth map generated by optimizing our energy function is shown in the third column, which is very similar to the ground truth for both diffuse and specular regions. Next, we remove the BRDF-invariant term, thus, only the gradient-based and smoothness terms are used for reconstruction ("Only Gradient"). The reconstructed depth map is shown in the fourth column, from which we can see that there is an obvious distortion caused by specular highlights. Quantitative comparisons in Table 1 also show higher errors when removing the BRDF-invariant term, which is consistent with intuition.

Next, we remove the adaptive weighting cue $\omega_{\mathbf{u}}(I, z)$ which balances between the BRDF-invariant term and the gradient-based photo-consistency term ("No Weighting"). The reconstruction again suffers from artifacts and fine details of object shape are oversmoothed. Similar effects can be observed from the surface reconstruction results by optimizing only the BRDF-invariant term and the smoothness term ("Only BRDF"). However, our surface reconstruction results without gradient-based photoconsistency term are still much better than those of [6]. This is because we use a better variational framework and a robust coarse-to-fine strategy to optimize the energy function. The reconstruction error using the two optimization frameworks are shown in the fifth and seventh columns of Table 1, respectively. Finally we use a Lambertian shading term to replace the BRDF invariance term for surface reconstruction, recreating the method of [2]. However, as shown in the seventh column of Figure 1 ("Diffuse Shading"), using Lambertian shading instead of our physically-based BRDF-invariance deteriorates the reconstruction results due to the highly specular surface, which is also observed quantitatively in Table 1.

2. Comparisons Using Real Data

Similar to Section 4.2 or Figure 7 of the main paper, we show surface reconstruction results with real light-field images using our method and compare with prior works, but using a depth map visualization.

The second column of Figure 2 presents the reconstruction using our method while the third column shows the results by removing the BRDF-invariant term. From the third column, we can see that minimizing the gradient-based stereo and smoothness terms can only recover most details for diffuse regions, but suffers from strong artifacts and distortions near the specular highlights. On the other hand, incorporating the BRDF-invariant term can effectively drive the optimization process, which leads to satisfactory surface reconstruction results for both diffuse and specular regions.

Next, we compare with methods such as [6, 4], which also use light-field cameras for surface reconstruction with unknown BRDF. As shown in the fourth column, even though our BRDF-invariant term is shared with [6], their surface reconstruction results are more noisy and inaccurate. This is because it fails to balance between BRDF-invariance and photoconsistency cues, while its optimization method is also not as robust as our variational approach. We also compare with a light-field reconstruction method based on a purely Lambertian assumption [5], in the sixth column,

which also exhibits strong distortions due to the glossy nature of the materials in these experiments.

3. Optimization Framework

For completeness, this section provides a detailed description of our variational framework for surface reconstruction. Following [2], we adopt the surface representation of [3] based on continuous bicubic patches and use a Gauss-Newton method for optimization. We next provide the derivation of the BRDF-invariant term in our energy function following [1, 6], except for details of camera setup and then formulate it with respect to our surface representation.

3.1. Surface Representation

We represent the surface using a set of square bicubic patches. The shape of each patch is controlled by four nodes placed on the image grid. Each node is represented by four values, the depth z, the first derivatives $\frac{\mathrm{d}z}{\mathrm{d}x}, \frac{\mathrm{d}z}{\mathrm{d}y}$ and the second derivative $\frac{\mathrm{d}^2 z}{\mathrm{d}x\mathrm{d}y}$. We optimize the four values of each node of the patch to recover the object surface. Let us denote the four nodes of a patch by 0, 1, 2 and 3. Let w be the size of the patch. Then, the normalized coordinate of a pixel covered by the patch is

$$\hat{\mathbf{u}} = (\hat{u}, \hat{v}) = (\frac{u - u^0}{w}, \frac{v - v^0}{w}) \tag{1}$$

The depth at each pixel is given by bicubic interpolation:

$$z(u, v) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij}(\hat{u})^i(\hat{v})^j \tag{2}$$

As a result, the first and second derivative of each pixel can be represented as

$$\begin{aligned} z_u(u, v) &= \frac{i}{w}\sum_{i=1}^{3}\sum_{j=0}^{3} a_{ij}(\hat{u})^{i-1}(\hat{v})^j \\ z_v(u, v) &= \frac{j}{w}\sum_{i=0}^{3}\sum_{j=1}^{3} a_{ij}(\hat{u})^{i}(\hat{v})^{j-1} \\ z_{uv}(u, v) &= \frac{ij}{w^2}\sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij}(\hat{u})^{i-1}(\hat{v})^{j-1} \end{aligned} \tag{3}$$

$$\begin{aligned} z_{uu}(u, v) &= \frac{i(i-1)}{w^2}\sum_{i=2}^{3}\sum_{j=0}^{3} a_{ij}(\hat{u})^{i-2}(\hat{v})^j \\ z_{vv}(u, v) &= \frac{j(j-1)}{w^2}\sum_{i=0}^{3}\sum_{j=2}^{3} a_{ij}(\hat{u})^{i}(\hat{v})^{j-2} \end{aligned}$$

The 16 parameters $\{a_{ij}\}$ can be calculated by solving a linear system once we know the values at the four nodes. Let $\mathbf{A}$ be a constant coefficient matrix derived from bicubic interpolation (2). By stacking the 16 parameters $\{a_{ij}\}$ into a vector $\tilde{\mathbf{a}}$ and the values of the four nodes into a vector $\tilde{\mathbf{z}}$, we have

$$\tilde{\mathbf{a}} = \mathbf{A}^{-1}\tilde{\mathbf{z}} \tag{4}$$

By combining (2), (3) and (4), we can easily compute $\frac{\mathrm{d}z}{\mathrm{d}\tilde{\mathbf{z}}}$, $\frac{\mathrm{d}z_u}{\mathrm{d}\tilde{\mathbf{z}}}$, $\frac{\mathrm{d}z_v}{\mathrm{d}\tilde{\mathbf{z}}}$, $\frac{\mathrm{d}z_{uv}}{\mathrm{d}\tilde{\mathbf{z}}}$, $\frac{\mathrm{d}z_{uu}}{\mathrm{d}\tilde{\mathbf{z}}}$ and $\frac{\mathrm{d}z_{vv}}{\mathrm{d}\tilde{\mathbf{z}}}$ using the chain rule. To minimize our energy function, we update the $\tilde{\mathbf{z}}$ vector for each patch iteratively using the Gauss-Newton method to obtain the surface reconstruction.

3.2. Derivation of BRDF-Invariant term

We follow the derivations of [6] to incorporate the BRDF-invariant term into our variational model. Here, we place the origin at the optical center instead of the center of the image plane, thus, the formulation is slightly different from [6]. As discussed in the main paper, the BRDF invariant energy term is defined as

$$E_{\text{BRDF}} = |(\kappa_1 + \kappa_2 z)n_x + (\kappa_3 + \kappa_4 z)n_y + (\kappa_5 + \kappa_6 z)n_z|, \tag{5}$$

where

$$\begin{aligned} \kappa_1 &= I_v H_{11} - I_u H_{12} \\ \kappa_2 &= H_{12}\beta(\gamma_2 + I_u\gamma_1) - H_{11}\beta(\gamma_3 + I_v\gamma_1) \\ \kappa_3 &= I_v H_{21} - I_u H_{22} \\ \kappa_4 &= H_{22}\beta(\gamma_2 + I_u\gamma_1) - H_{21}\beta(\gamma_3 + I_v\gamma_1) \\ \kappa_5 &= I_v H_{31} - I_u H_{32} \\ \kappa_6 &= H_{32}\beta(\gamma_2 + I_u\gamma_1) - H_{31}\beta(\gamma_3 + I_v\gamma_1). \end{aligned}$$

To compute the BRDF-invariant term, we need to compute the normal vector of each pixel, which will be introduced in the following.

Derivation of the normal vector Using the continuous surface representation based on bicubic interpolation (2), (3), (4), we have

$$z_u = \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} i (u)^{i-1}(v)^j \tag{6}$$

$$z_v = \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} u^i j (v)^{j-1} \tag{7}$$

With some abuse of notation, here (u, v) is the coordinate of the pixel on the image grid. Further, $\{\hat{a}_{ij}\}$ can be computed easily using the transformation $\hat{u} = \frac{u - u_0}{w}$ and $\hat{v} = \frac{v - v_0}{w}$ accordingly. Recall that $x = \frac{zu}{\bar{f}}$, $y = \frac{zv}{\bar{f}}$

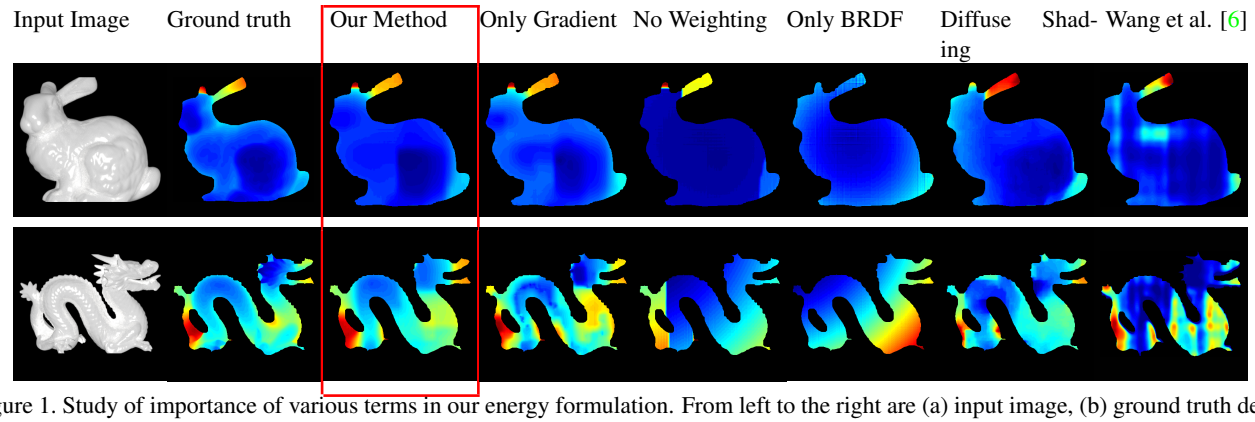

Figure 1. Study of importance of various terms in our energy formulation. From left to the right are (a) input image, (b) ground truth depth map, (c) depth reconstruction results by optimizing the full version of our energy function, (d) results without the BRDF-invariant term, (e) results without the weighting cue, (f) results without the gradient-based multiview stereo term, (g) results with the Lambertian shading cue of [2], instead of our physically-based BRDF-invariance and (h) BRDF-invariant term optimized using the framework of [6]. Please refer to the text for comparisons and analysis.

Model	Our method	Only Gradient	No Weighting	Only BRDF Term	Diffuse Shading [2]	Wang et al. [6]
bunny	**0.0011**	0.0014	0.0021	0.0258	0.0020	0.0551
dragon	**0.0025**	0.0032	0.0036	0.0116	0.0088	0.0194

Table 1. Study of the importance of various terms in our energy formulation. We use mean squared error to measure reconstruction accuracy.

where $\bar{f} = \frac{f}{s}$. Multiplying both sides of (7) by $\left(\frac{z}{f}\right)^6$, we obtain

$$\frac{z^7}{\bar{f}^6} = \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} x^i y^j (\frac{z}{\bar{f}})^{3-i} (\frac{z}{\bar{f}})^{3-j} the \tag{8}$$

Taking the derivative on both sides yields

$$\begin{aligned} 7\frac{z^6}{\bar{f}^6} z_x &= \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} i x^{i-1} y^j (\frac{z}{\bar{f}})^{3-i} (\frac{z}{\bar{f}})^{3-j} \\ &+ \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} x^i y^j (3-i) (\frac{z}{\bar{f}})^{2-i} (\frac{z}{\bar{f}})^{3-j} \\ &+ \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} x^i y^j (\frac{z}{\bar{f}})^{3-i} (3-j) (\frac{z}{\bar{f}})^{2-j}. \end{aligned}$$

Multiplying both sides by $\frac{\bar{f}^5}{z^5}$, we obtain

$$\begin{aligned} 7\frac{z}{\bar{f}} z_x &= \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} i x^{i-1} y^j (\frac{z}{\bar{f}})^{1-i} (\frac{z}{\bar{f}})^{-j} \\ &+ \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} x^i y^j (3-i) (\frac{z}{\bar{f}})^{-i} (\frac{z}{\bar{f}})^{-j} \\ &+ \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} x^i y^j (\frac{z}{\bar{f}})^{-i} (3-j) (\frac{z}{\bar{f}})^{-j} \end{aligned}$$

which may be written as

$$\begin{aligned} 7\frac{z}{\bar{f}} z_x &= \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} i u^{i-1} v^j + \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} u^i v^j (3-i) \\ &+ \frac{z_x}{\bar{f}} \sum_{i=0}^{3}\sum_{j=0}^{3} \hat{a}_{ij} u^i v^j (3-j) \\ &= z_u + \frac{z_x}{\bar{f}} (6z - u z_u - v z_v). \end{aligned}$$

Therefore, we have

$$z_x = \bar{f} \frac{z_u}{z + u z_u + v z_v}, \tag{9}$$

with a similar expression for the derivative with respect to y:

$$z_y = \bar{f} \frac{z_v}{z + u z_u + v z_v}. \tag{10}$$

In summary, the surface normal may be expressed as

$$\mathbf{n} = (z_u, z_v, -\frac{z + u z_u + v z_v}{\bar{f}}), \tag{11}$$

which may now be substituted into the expression for the BRDF-invariant term to be used in our energy minimization framework.

References

[1] M. Chandraker. The information available to a moving observer on shape with unknown, isotropic BRDFs. *PAMI*, 38(7):1283–1297, 2016. 2

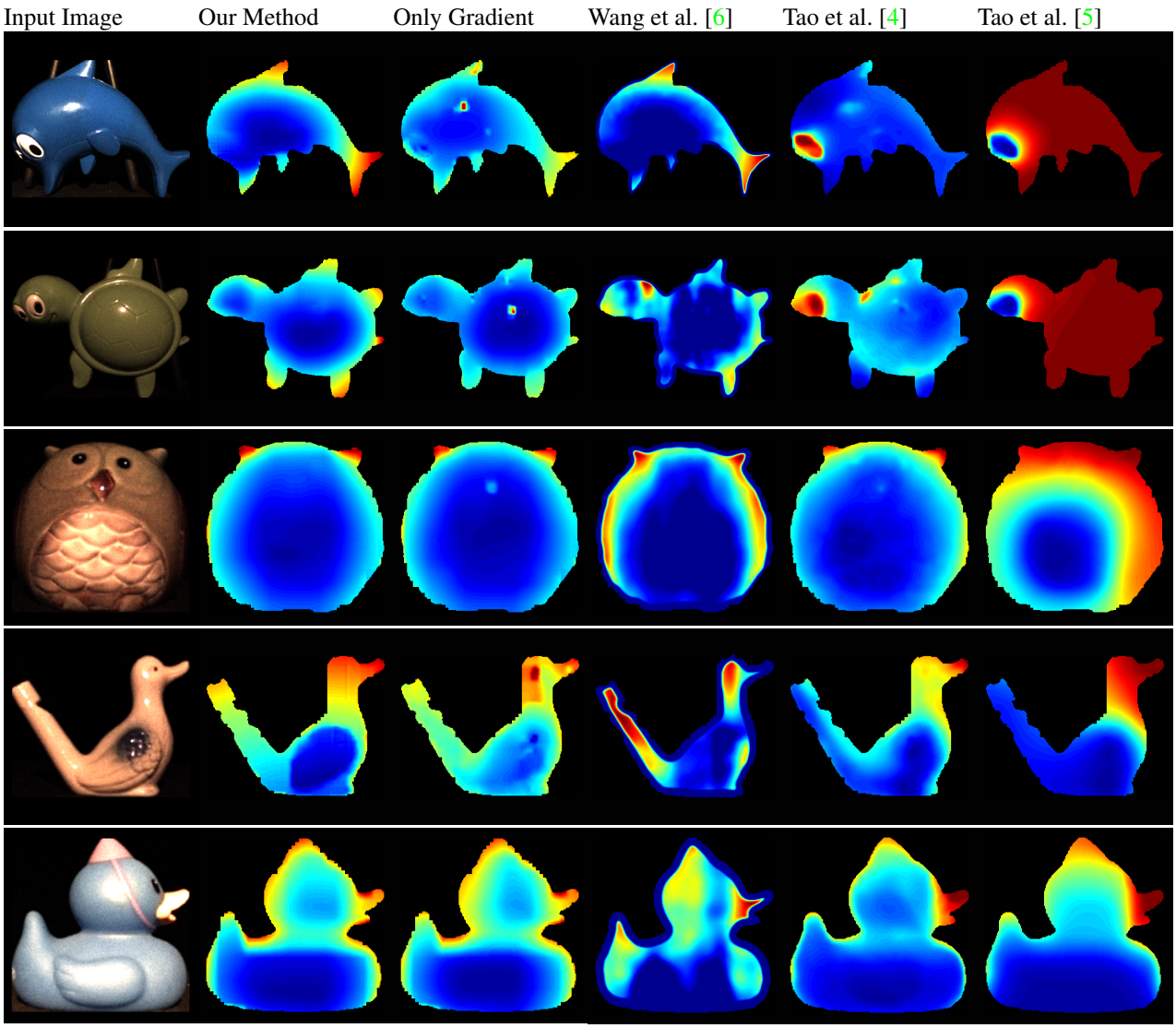

Figure 2. Single view shape reconstruction results on real data. From the left to the right are the original images, the ground-truth depth map, the reconstruction results using our method, the BRDF-invariant method of [6], the clustering method of [4] and the Lambertian method of [5]. We again observe that our robust optimization framework produces good reconstructions without sensitive parameter tuning.

[2] F. Langguth, K. Sunkavalli, S. Hadap, and M. Goesele. Shading-aware multi-view stereo. In *ECCV*, pages 469–485, 2016. 1, 2, 3

[3] B. Semerjian. A new variational framework for multiview surface reconstruction. In *ECCV*, pages 719–734, 2014. 2

[4] M. Tao, J.-C. Su, T.-C. Wang, J. Malik, and R. Ramamoorthi. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2015. 1, 4

[5] M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1940–1948. IEEE, 2015. 1, 4

[6] T.-C. Wang, M. Chandraker, A. A. Efros, and R. Ramamoorthi. SVBRDF-invariant shape and reflectance estimation from light-field cameras. In *CVPR*, pages 5451–5459, 2016. 1, 2, 3, 4